Issues in producing multi-platform CD-ROMs

Data Distribution Laboratory Staff

The following material is extracted from a review of Pre-mastering and CD-WO recorders performed by the Data Distribution Lab. The review will be published this spring in CD-ROM Professional Magazine.

Most CD-ROM titles produced to date have ignored the issue of multi-platform access to software and data. This is partly because such access has been difficult to provide without a great investment in time and effort by the producer. In developing the Mars Digital Image Mosaic CD-ROM collection, we decided to fully support PC, Macintosh, VAX and UNIX platforms. The production process involved:

1. receiving the CD-ROM data sets in VAX backup format on 8-mm tape;
2. restoring the files to a local VAX computer;
3. running a utility program that read the VAX directory for the volume and generated a CD-Publisher command file (including Extended Attribute Record 'XAR' specifications for VAX);
4. running the command file on the PC utilizing PC-NFS (PC - Network File System) to access the files on the VAX and build an ISO partition on the CD-Publisher system;
5. connecting the Publisher to a Macintosh and adding the Mac software directory to the ISO partition;
6. finally, generating a CD-ROM *image* file from the ISO partition.

Was the result worth the trouble? We can only say that there have been no complaints from any of the 500 recipients of the CD-ROM set and it is actively being used on all target platforms. Other issues in multi-platform CD-ROM production are described below.

Text files. Unfortunately, each of the major computer platforms has developed its own interpretation of what constitutes a simple text file. In the following examples <CR> stands for carriage return (hex 0D), <LF> stands for line feed (hex 0A) and <LENGTH> stands for a 2-byte length field stored in Least Significant Byte (LSB) first notation.

IBM PC	This is a line of text.<CR><LF>
Macintosh	This is a line of text.<CR>
UNIX workstation	This is a line of text.<LF>
VAX/VMS workstation	<LENGTH>This is a line of text.*

* This is the most common format supported under VMS.

Only the IBM PC format will display properly on the other platforms. For this reason the Data Distribution Laboratory uses the IBM PC version for all simple text files. There are vendor supplied utilities that can translate the IBM PC format to internal format. These are Apple File Exchange on the Macintosh, dos2unix or the translate (tr) utility on UNIX workstations and Convert on VAX/VMS workstations. The other option is to include four versions of each file, which can become a configuration management nightmare.

File naming. It is unfortunate that the limitations of MS-DOS have put stringent restrictions on file names that can be used on transportable CD-ROM volumes. The ISO 9660 standard

provides levels of interchange to insure that all file systems will be able to access CD-ROM files. Level 1 requires that names be composed of one to eight (upper case) characters with a three character extension. File names may only contain alphabetic and numeric characters and the underscore character. Other platforms are more lenient in the number of characters allowed (MAC - 31, VAX - 64, UNIX - 128) and which characters may be included. For example, Macintosh file names may include spaces. UNIX file names may contain a variety of special characters and multiple 'periods', which is disastrous on IBM PC or VAX systems.

File version numbers. The ISO-9660 standard requires a version number (abc.dat**;1**) on each filename (the dreaded semi-colon issue). Imagine building a Macintosh retrieval application that accesses files on the CD-ROM by filename. You test it thoroughly, then cut an ISO 9660 disk. Imagine your chagrin when you find your application does not work at all from the CD-ROM, because the pre-mastering software appended a ';1' version number to all file names. Different drivers for different operating systems handle the version number differently. The IBM PC ignores it. The Macintosh CD-ROM driver developers felt obligated to maintain the version number with the file name. Thousands of hours of effort have been expended (and wasted) due to this idiosyncrasy. Even worse, one cannot always assume that the version number will be 1. A CD-ROM pre-mastered on a VAX computer takes the version number from the original data file, so it will very often be a higher version number.

Macintosh Files. The Macintosh file system partitions files into resource forks and data forks. Resource information includes the file type and file creator, encoded as 4-byte identifying strings. Files are represented by the operating system as pictorial ICONs, which are also stored in the resource fork. If these files are pre-mastered on a standard ISO 9660 CD-ROM they will appear to the Macintosh as text files. Clicking on an application will not work. Thus Macintosh executable and special file types are most commonly stored in some compressed (typically Stuffit) or BINHEX format. They must then be uncompressed or un-BINHEX'ed to the user's hard disk before use. This approach has the advantage of avoiding the version number problem mentioned previously. It has the disadvantage of enormous user complexity. Try describing the process to a naive Mac user who does not have Un-Stuffit or Binhex on his/her computer.

Extended Attribute Records. Extended attribute records provide a method of retaining basic file characteristics (record type, record length, record attributes), file protection settings and time tags for files recorded on ISO CD-ROMs. These records are most important in the VAX/VMS environment, where record structure information is often required by software. The new VAX/VMS CD-ROM (System 5.5 and above) support allows default attributes to be set when a volume is mounted.

CD-ROM flavors

ISO 9660

ISO 9660 is the international standard for transportable CD-ROMs. Thanks to the support of Jerry McFaul's Special Interest Group for CD-ROM Applications and Technology (SIGCAT) every archival CD-ROM produced by government agencies has used the ISO standard. Figure 3 shows how the ISO 9660 file structure is organized using a sample CD-ROM, and how DOS interprets the file structure.

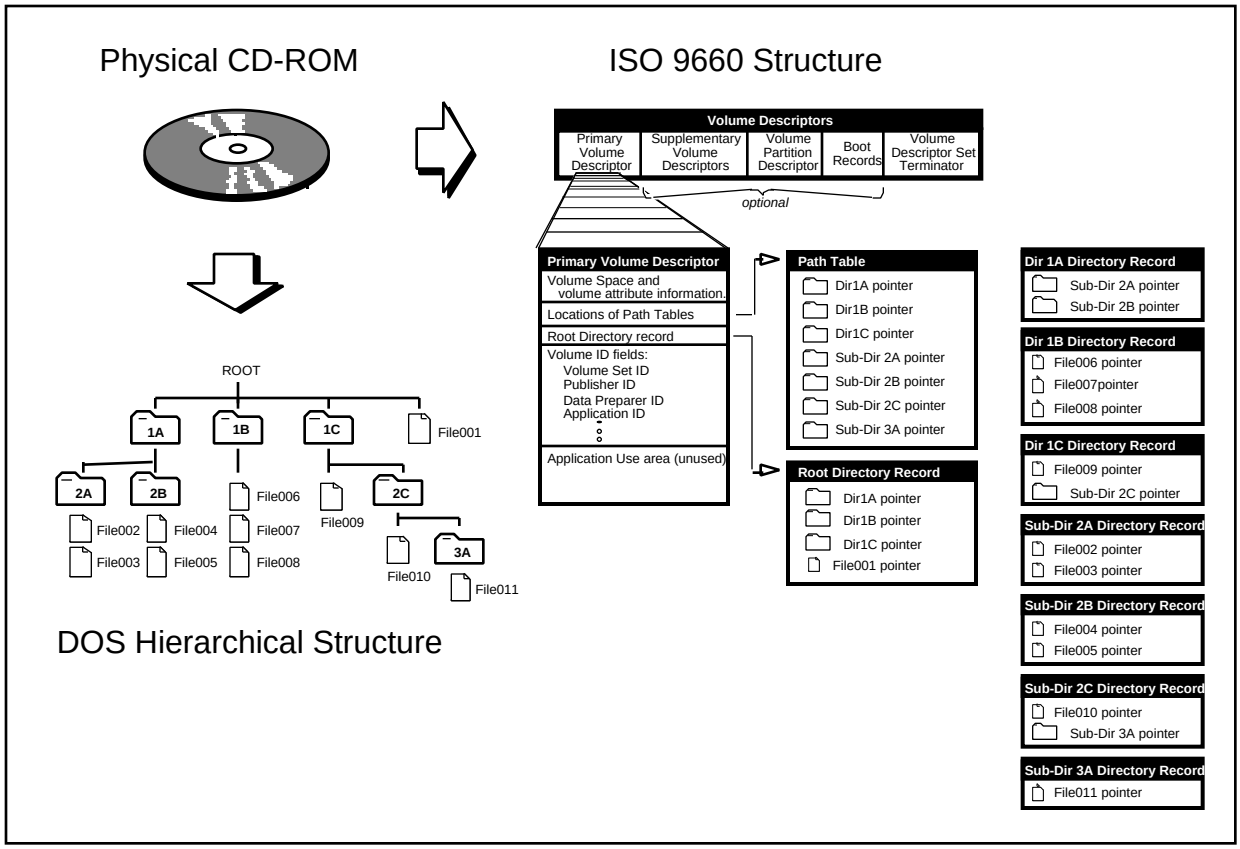

Figure 3: Example CD-ROM and how it is implemented in ISO 9660.

High Sierra

While this landmark 'standard' was useful for a period of about 9 months while ISO 9660 was being approved, High Sierra is NOT a standard and no one should use it! But, alas, Microsoft's new Bookshelf for Windows is still recorded in High Sierra format. Many PC directed titles use this format.

Macintosh Hierarchical File System (HFS)

Macintosh Hierarchical File System format is useful for making Macintosh-only CD-ROMs. Apple uses this format for its products, primarily for performance reasons. Methods of making Mac compatible ISO 9660 compliant CDs on other platforms are now available if transportability is an issue.

Macintosh ISO 9660 disk.

Apple has implemented a file system based on ISO 9660. Since the data and its information for a single file in HFS or MFS is stored in two different places called the Resource Fork and the Data Fork, extra features are required for mounting HFS/MFS files from ISO 9660 volumes. Thus, Macintosh implemented the System Use field in the directory record and the associated file, which is supported by the directory record file flags. Information that belongs in the Resource Fork is recorded in the associated file, while data resides in a normal file (see Figure 4). This implementation for the associated file is especially important for storing executable files, since all of an executable file and its information resides in the Resource Fork. However, in most cases, none of the non-Apple vendors implement these features for either pre-mastering

software or CD-ROM drivers. Thus, in order to record Apple Macintosh files, one must pre-master on Macintosh systems.

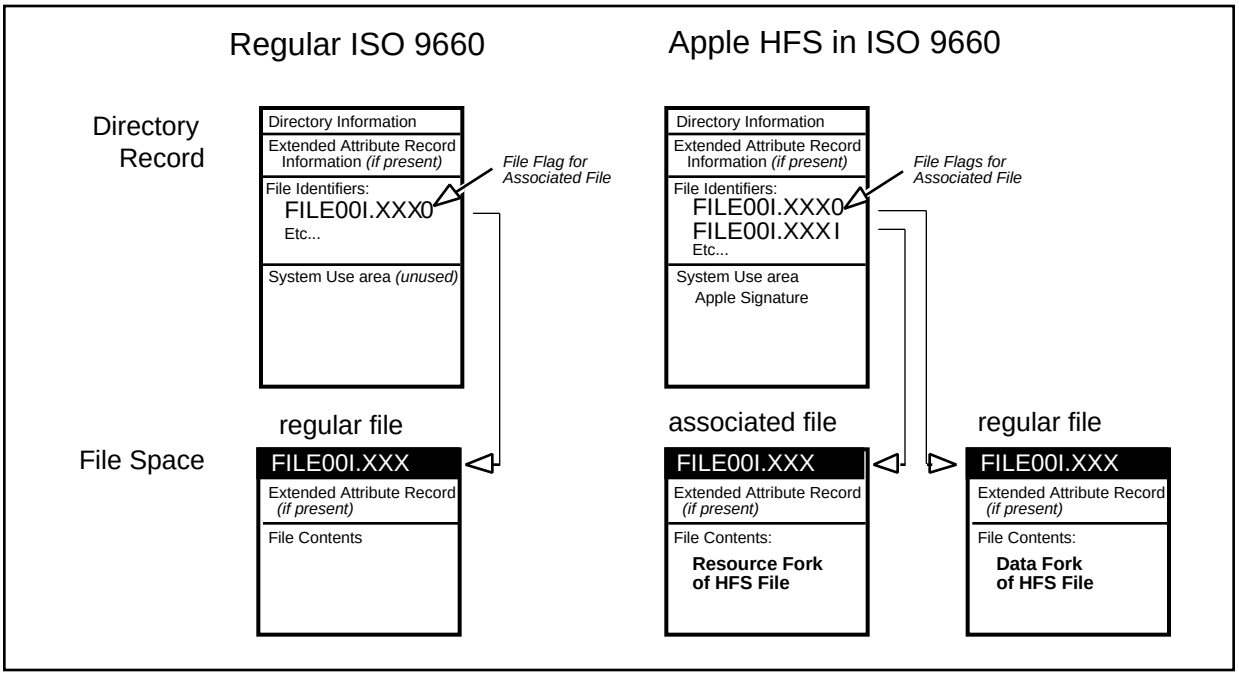

Figure 4: Implementation of Apple Macintosh HFS in ISO 9660.

Rock Ridge.

The purpose of the Rock Ridge initiative was to create and agree upon a common format by utilizing the system and application area in the directory record of the ISO 9660:1988 format while maintaining compatibility with the installed base of ISO 9660:1988 hardware and software in the UNIX environment (see Figure 5). This proposed standard uses the System Use Area and Application Use Area in the directory record to include necessary information such as UID, GID, mode bits, major and minor device, UID, and GID numbers by receiving systems for X/Open systems. Thus, this scheme will eliminate the XAR fields and increase CD-ROM drive performance. It also sets up a standard procedure for sharing System Use Areas.

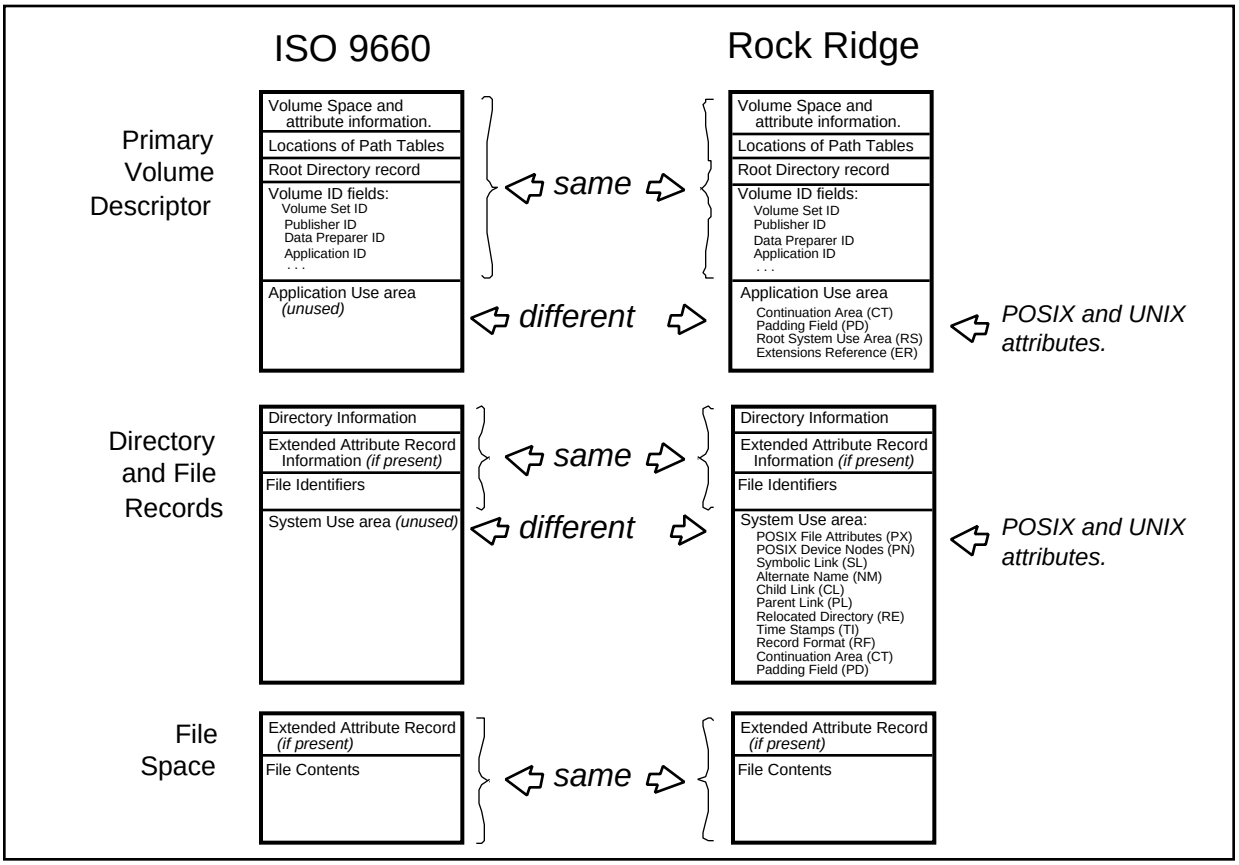

Figure 5: Differences between ISO 9660 standard and Rock Ridge Proposal.

CD-WO Standard

CD-WO is an evolution of CD-ROM to a sequential write-once medium and is defined by the Orange Book. A group called the ECMA (European Computer Manufacturer Association) TC15 working group, informally known as the Frankfurt group, is working on the volume and file structure standard for CD-WO. This standard group has produced ECMA 167, which defines volume and file structure for WORM and Erasable media, and ECMA 168, which is upward compatible with the ECMA 119 and the ISO 9660 standards. These two ECMA standards are expected to be accepted as ISO standards.

UNIFILE

UNIFILE is Digital Equipment Corporation's (DEC) CD-ROM format, used on most of DEC's VAX/VMS CD-ROM products. With Version 5.5 of VAX/VMS, DEC now provides built-in supports for ISO 9660 and High Sierra CD-ROMs, and hopefully UNIFILE will disappear.